What is to be explained?

Jeremy Attard*

December 5, 2022

Abstract

The deductive nomological (DN) model has been the basis for discussions about scientific explanations for decades. The overcomming of the logical empiricist program together with the raise of several counter-examples to the DN model have progressively led to a renewal of the reflections on this topic. The first step of this paper is to clarify the framework in which the epistemological question of scientific explanation is adressed. We make a proposal for a universal structure of scientific models, which constitute the basic epistemic unit of our analysis. It is then possible, within this framework, to clarify some discussions about scientific explanation, and to offer a new account of it. The latter tries to benefit from the advantages of the DN model, as resting on a law, together with neutralizing some of its typical criticisms. The work presented here is both abstract, exhibiting formal structures as general patterns of explanation, and concrete with real examples taken from actual scientific disciplines.

Keywords: Explanation – Scientific Model – DN Model – Analytical Sociology.

*Philosophy and History of Science department, University of Mons (Mons, Belgium) & Department of Sciences, Philosophies and Societes, University of Namur (Namur, Belgium).

1 Introduction

Modern discussions about scientific explanation have been mostly structured around the deductive nomological (DN) model due to Carl Hempel and Paul Oppenheim (Hempel & Oppenheim 1948), (Hempel 1965). In its simplest and general form, and as its very name suggests, the DN model claims that explaining means deriving from a law. More precisely, an explanation has the following general structure:

$$\text{Explanans} \xrightarrow{\text{Logically entails}} \text{Explanandum}$$

Broadly, the *explanandum* denotes the target of the explanation, i.e. what is to be explained, and the *explanans* is what explains. For instance, in the statement "The road is wet because it has rained", the road's wetness is the explanandum, and the rain is the explanans. The DN model for scientific explanations goes further and states that the explanans must contain some general laws which are necessary to derive the explanandum. For instance, the observation "the stone felt down to the ground" is explained by the fact that "I hold the stone and then I dropped it" *if* the general law "when material objects of a density greater than air's are dropped, they fall down" is hold to be true.

The DN model has been very inspiring for these discussions about scientific explanations over years, yet nowadays it seems to have been abandoned in its original form. (See e.g. (Salmon 1989), (Barberousse et al. 2011, chapter 1) and (Woodward & Ross 2021) for a comprehensive review.) The DN model dates back in the time when logical empiricism (in its diverse manifestations) was the main epistemological paradigm, and it turns out that the DN model is well suited for this paradigm which sees scientific theories as collection of statements with axiomatic structures. Moreover, a certain number of counter-examples to the DN model have raised over time and seem to show that the DN structure (and more precisely the presence of a law) is neither sufficient nor necessary to be a scientific explanation.

In the field of social sciences and more precisely that of sociology, the very question of explanation as a relevant aspect or aim of the discipline has been adressed from its very begining. Max Weber's 1922 *Gesammelte Aufsätze zur Wissenschaftslehre* (Weber 1922, 1965) is an attempt to solve the conflict of methods (*Methodenstreit*) which lasted for the end of the nineteenth century between *understanding* and *explanation* as two possible and rival aims of sociology. This distinction broadly corresponds, respectively, to the view which claims that sociology (and social sciences in general) are of a distinct kind w.r.t. other sciences (and especially physics) and to the view which claims that all sciences can (and must) be understood under the same epistemological paradigm, despite of their great diversity of objects and empirical methods. Nowadays, sociology is still a wide and heterogeneous field of research both from scientific and epistemological viewpoints. It goes for very descriptive, micro scale approaches without any pretention to generality to attempts to describe general social phenomena within mathematical models, passing by participant observations, the use of statistics as well as even direct experiments. It is not our aim here to draw a precise and historical description of the current landscape of the social sciences in all their diversity, neither to summarize the different epistemological viewpoints which have existed and still co-exist today. We come back quickly to this discussion later on. Our point here is merely to notice that a direct application of the DN model to the social sciences, and more precisely to sociology, is all but straightforward - and raise in particular the question of the existence of laws in the social world. This is thus an additional issue which questions the relevance of the DN account of scientific explanation.

A first useful distinction for our discussion is the one between the *structure* of a scientific explanation and its epistemological *value*. The first question to ask is indeed: what something claiming at being a scientific explanation looks like? What is the *structure* of such a (claiming)-explanation? The second question is then: among different propositions for the scientific explanation of a given phenomenon, how can we recognize a good scientific explanation? How can we assess the *value* of such explanations in order to compare them? The answer to the second question necessary depends on the answer to the first one. Indeed, in order to compare some epistemic units betwen them, they have to share some common features. That is to say, good scientific explanations and bad scientific explanations, to be judged so relatively to each other, must share at least a minimal structure.

In the literature, both features are often confused when it comes to give a set of criteria characterizing a good scientific explanation. For instance, consider the original DN model (Hempel & Oppenheim 1948): the explanandum has to be a logical consequence of the explanans, and the latter must contain some laws which are necessary to derive the explanandum, and the statements made in the explanans have to be true. Here, both kinds of features (about structure and about value) are mixed. Indeed, the fact that the explanandum must be a logical consequence of the explanans and that the latter contain laws are *structural* constraints. Then, between different epistemic units aiming at explaining the same explanandum and having the same structure, we can hierarchize them in terms of their respective *value* based on the fact that the law in the explanans must be logically necessary and that the statements in the explanans have to be true.

In this paper, both aspects are alternatively under study: we first focus on structural aspects and then adress the question of a good or bad scientific explanation. We choose scientific *models* as the relevant part of analysis. Our work is then divided in two parts. First, we precisely define the epistemic unit under study, presenting a general schema of the structure of scientific models aiming at being universal. We do not (and actually cannot) demonstrate this universality on a definitive way. However, we support our arguments with examples in physics, sociology and epidemiology. Indeed, concerning sociology (and the social sciences in general), we take the decision to study a specific theoretical frame, that of *analytical sociology.* We relate our general meta-model of scientific models with the Coleman diagram used in this field, and also show the connection with the Bradford Hill's criteria in epidemiology. Once this general structure is introduced,[1]

[1]Again, from our viewpoint, exhibiting this structure is necessary to be scientifically judged as an explanation, but it is not sufficient in order to be a *good* explanation.

it allows in a second part to clarify some discussions about scientific explanations. In this aim, our strategy is to adress the question of what is a good *explanandum.* Indeed, we do not focus on what is a good explanans but rather on how to detect an inadequate explanandum, i.e. something which we may want to explain whereas there is actually nothing to be explained. From the general meta-model of scientific models, we then present three typical kinds of such inadequate explananda, with concrete examples from actual science. As we show in this paper, it allows to integrate some benefits of the DN model while neutralizing some classical counter-examples.

2 A universal structure for scientific models

A first step in this attempt to assess the quality of an explanation which claims at being scientific is to defined precisely the epistemic unit under study. The DN model originally conveniently rested on a logical empiricist viewpoint, thus the overtaking of this program calls for a new model to embody epistemological discussions in general, and discussions about scientific explanation in particular.

Our approach takes scientific (or scientific-claiming) *models* as privileged unit of analysis. We defend the idea that behind their great diversity, scientific models share a common structure that we aim at enlightening in this section, which is divided in three parts. First, we present a general (meta-)model for the structure of scientific models. Some concrete examples from physics will illustrate our point. Second, we take the case of the social sciences, and in particular that of analytical sociology to show the particular form that this (meta-)model takes in this field, in connection with the well-known Coleman diagram. Finally, we enlighten the strong analogy between this viewpoint and the Bradford Hill's criteria in epidemiology and biomedicine.

2.1 The structure of scientific models

2.1.1 General presentation

Figure 1 sketches our proposal for a general structure of scientific models, i.e. the epistemic unit underlying our whole current analysis.

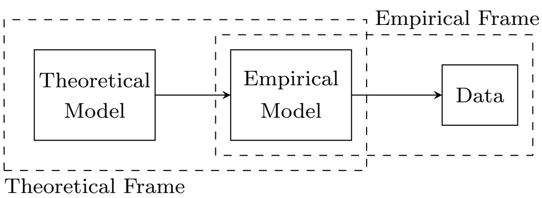

Figure 1: The general structure of scientific models.

This meta-model is composed of different parts.

- **The empirical frame.** It consists in a set of basic variables and magnitudes, together with their operationalized definitions. That is, sets of concrete operations necessary to give these variables a value. These variables can be of different nature: quantitative (continuous or discrete), qualitative (nominal or ordinal), either time-dependent or not, etc.

- **Data.** The empirical frame defines the empirical language at use to talk about a part of the world, i.e. about a concrete empirical situation. A fundamental ingredient is obviously empirical data, which can take a great diversity of forms depending on the objects under study and the variables choosen.

- **Empirical model.** An empirical model is defined here as a set of general claims aiming at recovering some empirical data, or at least which can be compared to some (virtual or actual) data. It is the form an empirical fact takes.

- **Theoretical model.** A theoretical model is composed of two parts: 1/ a representational part, which aims at representing the target, i.e. the empirical situation (the system, its environment, etc.) under study. 2/ An explicative part, which aims at deriving some empirical facts from (theoretical)[2] modelization hypothesis.

- **Theoretical frame.** It is composed of a set of general statements defining the frame within which the explicative model is written. That is, for instance, the fundamental principles defining the rules an explicative model has to follow.

Dashed rectangles in the figure 1 represent theoretical and empirical frames. The empirical frame is made explicit to remind the well-accepted epistemological fact that experience is always *theory-laden*. That is to say, there is no such things as pure observations: data are always generated within a certain frame and partly dependent on it.[3] The theoretical frame is also made explicit to remind the same kind of observations about the theoretical models: they rest on some assumptions (modelization hypothesis, fundamental principles, ...) and are not even meaningful outside of this frame.[4] We draw them overlapping to emphasize the fact that empirical models are written in a language which is common both to theoretical models and to the data. However, the both frames are not totally identified. On the first hand, some theoretical terms (like fundamental principles, or some modelization hypothesis) do not have a direct empirical meaning. On the other hand, some empirical terms entering into account in a given model are defined independently to the theoretical frame of this model.

In a common quinean view (Quine 1951), we see our whole knowledge as a vast logical web of statements the structure of which is eminently complex and interdependent.

[2]That is to say, hypothesis which do not necessarily have a direct empirical meaning.

[3]An experiment in physics is made using experimental devices the functioning of which rest on theories (identical or different from the one which is tested), i.e. on some assumptions that we need to hold as true to be able to get our data meaningful. In neuroscience, an image of the brain activity is not direct neither, for it consists mainly in detecting oxygen, under the assumption that neuronal activity is somehow related to oxygen flux in the brain.

[4]The gravitational force $\mathbf{F} = \frac{GmM}{r^3}\mathbf{r}$ has no specific meaning outside of the theoretical frame of Newtonian physics, i.e. its mathematical modelization decisions and conventions together with its fundamental principles.

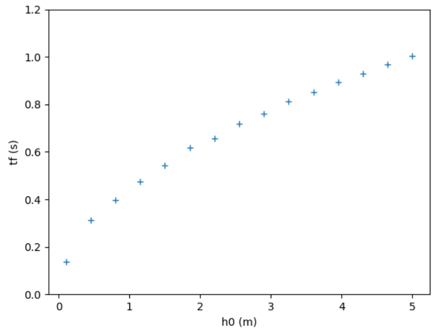

Figure 2: Data: fall time t_f of a spherical marble in vaccum as a function of the initial height h_0.

Defining such a unit of analysis means choosing to work at a certain scale (here, that of models). This meta-model has the same methodological use as that of models in science: simplifying the problem under study in order to end up with a cognitive construct easier to handle aiming at clarifying the discussion and which stays open at improvements. Choosing a certain (restricted) scale of analysis erases some details and we should remind that hypothesis constituing both the empirical and theoretical frames are connected to other hypothesis and models outside or inside the same frameworks. Duhem-Quine holistic thesis is thus not ignored here, but we assume that working at this precise scale is an approximation which is good enough for our purpose.

2.1.2 Illustration: free fall

A simple illustrative model is that of the fall of a spherical marble in vaccum. We take a spherical marble and we let it fall from an initial height h_0 from the ground. We then measure its fall time t_f and plot it w.r.t. h_0. The **data** we get look like in figure 2.

The corresponding **empirical frame** here is composed of height z and time t in general which then are adapted to measure initial height h_0 and final time t_f in particular.

A possible **empirical model** of this phenomenon is the following:

$$t_f = \sqrt{K.h_0} \text{ with } K \approx 0.2 \text{ m}^{-1}\text{s}^2 \tag{1}$$

Then, a **theoretical model** has to derive this empirical regularity from theoretical considerations. A straightforward way to do that is to work within the framework of Newtonian mechanics which thus constitutes the **theoretical frame** of the model. The marble is modelized by a material point of mass m in a three dimensional euclidean space of basis $(\mathbf{e_x}, \mathbf{e_y}, \mathbf{e_z})$ the position of which is represented by a vector $\mathbf{r}(t) = (x(t), y(t), z(t))$ in this basis. To his material point, a finite number of forces $\{\mathbf{F}_i\}_i$ is assumed to apply. The fundamental principle of Newtonian dynamics states that:

$$m\ddot{\mathbf{r}} = \sum_i \mathbf{F}_i \tag{2}$$

where $\ddot{\mathbf{r}}$ is the second derivative of the position w.r.t. time (i.e. the acceleration). The theoretical frame also assumes that in the case of a fall in vaccum (i.e. there suppose to be no fluid in which the marble falls), the only force which applies is the Earth's gravitational force:

$$\mathbf{F} = m\mathbf{g}. \tag{3}$$

$\mathbf{g}$ is the Earth's gravity field which reads: $\mathbf{g} = -g\mathbf{e_z}$ with $g \approx 10 \text{ ms}^{-2}$. From equations (2) and (3), together with the initial conditions: $\mathbf{r}(0) = h_0$ and $\dot{\mathbf{r}}(0) = \mathbf{0}$, we get: $\ddot{z}(t) = -g$ and then:

$$z(t) = h_0 - \frac{1}{2}gt^2. \tag{4}$$

Fall time t_f is such that $z(t_f) = 0$. Once plugged into formula (4), we get

$$t_f = \left(\frac{2h_0}{g}\right)^{1/2} \tag{5}$$

which gives back the empirical model (1) with $K = 2/g$. The overlapping of the empirical and theoretical frames, as drawn in figure 1, is represented by the fact that the variables z and t have both an empirical meaning (the way they are indeed measured in the concrete empirical situation) and a theoretical one (their respective role in the theoretical model). The both frames still do not identify since some terms, like Newtonian forces $\{\mathbf{F}_i\}_i$ do not have a direct empirical meaning: they appear as fundamental ingredients of the theoretical model, but eventually disappear.

2.2 The Coleman diagram and the social sciences

2.2.1 Analytical sociology

The human and social sciences, particularly sociology, form a wide and heterogeneous intellectual landscape. Methodological, epistemological and scientific debates about fundamental issues are still vivid. For our purpose, however, we choose to restrict our analysis to a quite recently institutionalized branch of sociology, namely analytical sociology. Analytical sociology has been formed as a research program in the 1990's on the heritage of a certain kind of sociology as promoted by (among others) Robert Merton, James Coleman or Raymond Boudon. See e.g. (Hedtröm & Bearman 2009), (Manzo 2014) or (Manzo 2021) for the most recent states of the art of the discipline. From that time, this field has grown, structured and have been inhabited by constant epistemological discussions. Broadly, analytical sociology walks on two fundamental legs. On the first hand, a call for conceptual clarity, well-defined quantities and entities, and analytical rigor. On the other hand, one of its central aim is to base the description and the explanation of social phenomena on clear, transparent and plausible *social mechanisms* at the microsocial level aiming at explaining empirical regularities observed at the macrosocial one.

2.2.2 The Coleman diagram

A well known way of representing an important part of this research program is what is usually called the Coleman diagram (or Coleman boat), see figure 3 (Coleman 1990), (Ylikoski 2021).

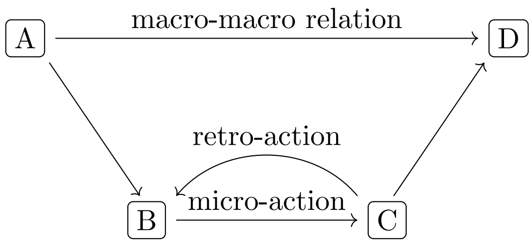

Figure 3: The Coleman diagram

This diagram is composed of the following parts:

- $A \to D$ represents a relationship between two variables at the macro level. It is not a mere statistical observed correlation between two variables, but a genuine inductive hypothesis, as e.g. a hypothetical causal relationship between these two variables.
- $A \to B$ represents the (causal) influence the macro state A has on the micro state B. B represents for instance the limited set of possible choices an agent can have, and this set of possibilities is conditioned by some social states at the macro level.
- $B \to C$ represents how a certain micro state will influence another one. For instance, how the limited set of possibilities given to an agent influences the actions of the latter. There can also be some retro-action, e.g. in the case where many agents are interacting: the set of possible choice (and its corresponding probability distribution) of a given agent then depends on macro state A *and* on the action of other agents. This is a possible way to modelize social interactions, which must rest on a clearly exhibited mechanism.
- Finally, $C \to D$ represents how the actions of agents at the micro level aggregate and eventually produce the macro state D.

This research program is thus not holistic, in the sense that explaining macro facts necessitates to decompose social reality in smaller parts at a lower level, the latter's being analysed through mechanisms-based models (arrow $B \to C$). Then, a genuine explanation is obtained when the second macro state can be recovered from this micro-analysis (arrow $C \to D$), while also taking into account the influence of the macro level on micro actions (arrow $A \to B$). Therefore, this way of explaining social phenomena does not reduce to a naive atomism neither.

From this viewpoint, the arrow $B \to C$ is the core of the explanation. At the micro level, this has to be embodied into a clear (and plausible) mechanism. This mechanism is composed of some modelization hypothesis which are assembled with each other using a certain theory of (human) action.

2.2.3 Connection with our meta-model

From our viewpoint, this form of explanation in the social sciences has nothing deeply distinct from the one used in other sciences, especially in physics. We claim that the Coleman diagram (figure 3) is a particular case of the more general meta-model presented in section 2.1.1 (figure 1).

A first step towards this identification of both viewpoints is shown figure 4.

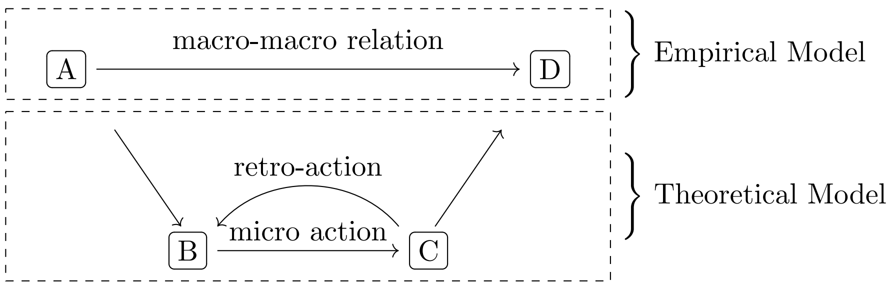

Figure 4: First correspondance between the meta-model and the Coleman diagram.

The equivalent of the empirical model is the macro-macro relation between both social states A and D. This is an empirical model because it goes beyond the data (if there are some) and assume a certain relationship at this level (as a causal relationship, going beyond merely observed correlations). Then, the theoretical model has to account for how micro states are influenced by the macro ones, which mechanism is at work at the micro level, and then how the aggregation of what is going on at the micro level produces the observed macro state. This theoretical model then relies on some theoretical frame as well: the mechanism at work rests on some theoretical and general assumptions (a theory of human action), which the mechanism is a particular instance of.

A second and last step to finalize the identification of viewpoints is to add a representation of empirical frame and data, which are not present in figure 4. In the social sciences, an example of such an empirical frame could be a finite[5] set of variables $\{v_1, v_2, ...\}$ (again, possibly of different kinds). An example of data could then be the observation of significant relationships between these variables in different configurations.[6] The force of these relationships can be quantified, e.g. with the correlation coefficient (for quantitative and continuous variables), the odd-ratio (for categorical variables), and so on – that is to say, the *size effect* of the relationship. Let α_{ij} denote the size effect measured between variable v_i and v_j. A possible schematic representation of an empirical frame and of data is given figure 5.

[5]The fact that the set of variables is finite is *per se* a hypothesis, for it states a certain number of factors which are assumed to be relevant for the purpose.

[6]For instance, observation of correlations between two variables once the other potentially relevant ones are fixed - i.e. *ceteris paribus*.

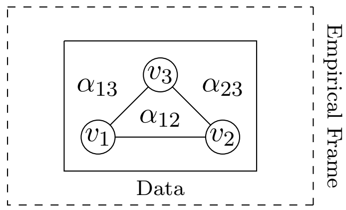

Figure 5: Empirical frame and data

Finally, we can propose a representation of the Coleman diagram as a particular instance of our meta-model in figure 6.

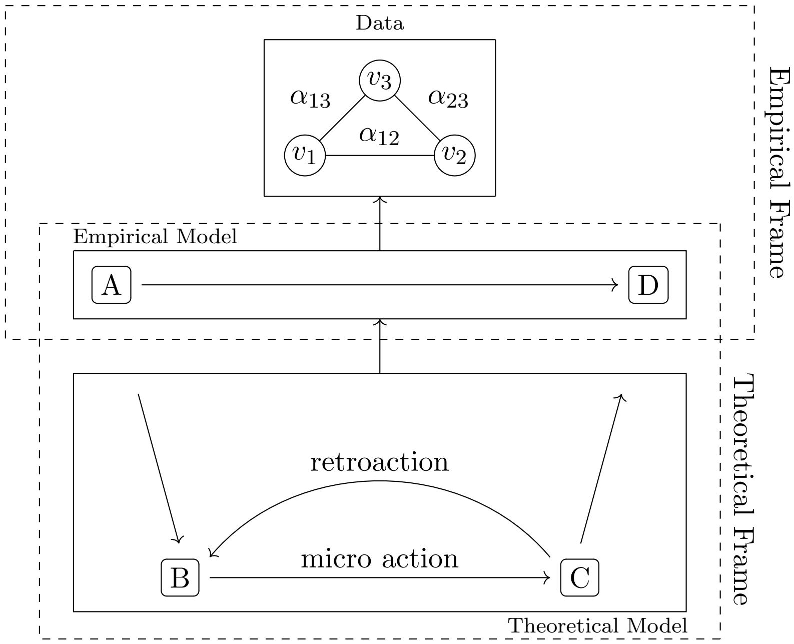

Figure 6: Meta-model of scientific models in the social sciences.

In the next section, we embody this abstract represention with a classical illustration: Max Weber's account for the rise of modern capitalism.

2.2.4 Illustration: Max Weber's account for the rise of modern capitalism.

We present here the example given by James Coleman himself in order to present his diagram: Max Weber's account for the rise of modern capitalism (Weber 2001). The first stage of his work is at the macro level: he finds a strong statistical association between the development of protestant values in a given society and their degree of capitalist organization. More precisely, Weber notices that (Weber 2001, p. 3):

> [statistics] indicate that people who own capital, employers, more highly educated skilled workers, and more highly trained technical or business person-

> nel in modern companies tend to be, with striking frequency, overwhelmingly Protestant. [...] According to the statistics, this variation between Catholics and Protestants is prominent where differences in religious belief and in nationality are found in the same region (and hence differences in the extent of historical development). Germans and Poles in northern central Europe come to mind. Yet the numbers demonstrate as well that differences are equally apparent in nearly all areas where capitalism, in the period of its great expansion [in the eighteenth and nineteenth centuries], possessed a free hand to reorganize a population. As these transformations took place, populations then changed according to indigenous paths of development, both socially and occupationally. In the process, differences according to religious belief became all the more striking.

In the footnotes of the first chapter of his book, he precises even more these statistical associations, emphasing that these associations remain significant not only within a comparison between nations but also between regions in the same nation or between some groups in the same religion. Other empirical regularities are given by a strong association between the students' choices about noncompulsory schools and their religion, on the basis on his student Martin Offenbach's work (Offenbacher 1901).

These are the **data**. The corresponding **empirical frame** is the set of all variables at use (and their definitions): religion, taxable wealth, countries/regions, choice of attending (or not) compulsory schools, education level, and so on. Here it is made clear that variables can be either quantitative (taxable wealth) or qualitative (religion).

The **empirical model** in this work assumed the existence of a causality relationship between the adoption of Protestant's (and more precisely Calvinist's) ethics within a certain region (e.g. in a country) and the development of capitalist economical organization. The **theoretical model** which aims at explaining this macro level assumption is summarized by Coleman (Coleman 1990, Chapter 1, p. 8) as follows:

- Protestant religious doctrine generates certain values in its adherents. [Arrow $A \to B$ in figure 6.]
- Individuals with certain values [...] adopt certain kinds of orientations to economic behavior. [...] [Arrow $B \to C$ in figure 6.)]
- Certain orientations to economic behavior [...] on the part of individuals help bring about capitalist economic organization in a society. [Arrow $C \to D$ in figure 6.]

The mechanism is thus the following: the religious doctrine (Calvinism) generates certain values in individuals (antitraditionalism and duty to their calling) which then are more prompt to adopt behaviors (labor and accumulation) which in turn facilitate the rise of capitalist mode of production.

What about the **theoretical frame** at work here? This mechanism necessarily relies on some theoretical assumptions or principles, for instance the assumption that an individual tends to behave according to her values (principle of axiological rationality). If this principle is not assumed to hold, then the theoretical model just looses its explicative power.

2.3 Connection with Bradford Hill's criteria

In the field of epidemiology and biomedicine, it is not always possible to conduct randomized controlled trials (RCT) in order to test a hypothetical causal relationship, for instance between a food habit and a cancer risk. What is available are, at best, observational data, e.g. measures of a certain set of variables on some statistical samples of the target population. Some statistical associations can be observed, quantified e.g. by the size effect. Quite strong statistical associations may suggest an underlying causation, but characterizing the direction of this causation and the corresponding variables in that are involved is anything but straightforward.

In 1965, Austin Bradford Hill proposed a set of criteria (Bradford Hill 1965) in order to assess to which extent it is reasonable to assume a causal relationship behind the observation of some statistical associations. His criteria are not assumed to be neither sufficient or necessary, that is why it is more common to talk about Bradford Hill's *guidelines for causation*. These guidelines have been extensively discussed over the past decades. We base our current discussion upon a quite recent revision of these guidelines (Howick et al. 2009). The authors organize these criteria into three new categories of evidence:

- Direct evidence: observed statistical associations with plausible confounders ruled out.
- Mechanistic evidence: existence of a plausible (e.g. biological) mechanism.
- Parallel evidence: existence of related studies with similar results.

These criteria are, to some extent, proper to the field of epidemiology or biomedicine. However, the categorization of types of evidence fits well with the general structure of scientific models proposed in section 2.1.1, and more precisely with the version of Coleman diagram (see figure 6) given in section 2.2.3. Indeed, *direct evidence* category corresponds to data and empirical model in our framework. We defined an empirical model as going beyond the data, e.g. alleging some causal relationship behind an observed statistical association. It still rests on some data in the sense that this causal hypothesis do not only have to reproduce the observed association, but also the fact that the association remains strong even when other relevant variables (confounders) are controlled. *Mechanistic evidence* category corresponds to the explanatory (theoretical) model which describes what is going on at the micro level. The only difference is that the term "micro" refers to individuals (for instance) in the case of sociology and to biological or chimical processes in biomedicine. Yet, this last remark does not have any incidence on the relevance of our point: the structure of a model is the same in both cases. Finally, *parallel evidence* category does not bear directly on the model's structure but rather on the expected quality of the evidence, so has no direct corresponding in our model since the latter focus on the structure of scientific explanations and not on its epistemological value.

3 Some inadequate *explananda*

The purpose of this section is to use the framework developed in the previous one, mainly focused on the general structure of scientific models, to clarify some discussions about scientific explanation, both on the structural side (what is the structure of something claiming to be a scientific explanation?) and on that of the epistemological value (how to distinguish between (relatively) good and bad explanations?). The first part is dedicated to using the general schema of scientific models to enlighten the general structure of scientific explanations. It turns out that two possible structures can be exhibited: practical explanations (where the explanandum is the data and the explanans is the empirical model) and fundamental explanations (in which the empirical model turns into the explanandum, and the explanans is the theoretical one). From these considerations, some preliminary clarifications are made in the second part, especially concerning the relationship between laws, explanation and prediction. Finally, in a third part, we adress the problem of what is a good explanation by focusing on the explanandum. Our strategy is to exhibit three kinds of inadequate explananda, i.e. claims that seems to call for an explanation while there is actually nothing to be explained. These reflections are examplified on some classical counter-examples of the DN model and eventually neutralize them in a quite natural way.

3.1 Scientific *explandum* and *explanans*. Two levels of explanations.

Let us starting again from our meta-model, with simplified labels:

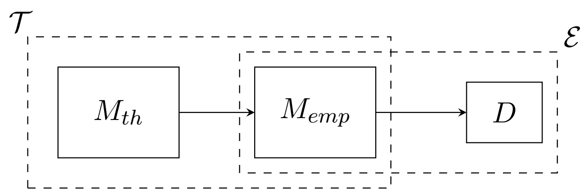

Figure 7: Meta-model of scientific models. $\mathcal{T}$ and $\mathcal{E}$ stand for the theoretical frame and the empirical frame, respectively. M_{th}, M_{emp} and D stand for the theoretical model, the empirical model and the data, respectively.

Let $EXPM$ stand for *explanandum* and EXPS stand for *explanans*. As recalled in the introduction, $EXPM$ is what is to be explained, and $EXPS$ is what explains. We can schematize their logical relation as:

$$EXPS \xrightarrow{\text{Logically entails}} EXPM$$

This is the general schema for explanations in the broad sense of the term. Now, we

Practical Explanation

$$(\mathcal{E}, M_{emp}) \longrightarrow D$$

Fundamental Explanation

$$(\mathcal{T}, M_{th}) \longrightarrow M_{emp}$$

Figure 8: Two legitimate levels of scientific explanations

restrict our work to *scientific* explanations: the first step is then to state the correspondance between $EXPS$, $EXPM$ and the parts of the meta-model figure 7.

The correspondance is not straightforward for there are actually two ways of understanding scientific explanations. Indeed, consider the following reasoning:

- Alice: *Why did the stone fall down?*
- Bob: *Because I hold it in my hand and then I dropped it.*

What is implied by Bob is that the following general statement: *When stones are held at a certain height and then dropped, they fall* is true. This is the typical structure of the DN model of explanation. The "law" here is the empirical regularity, and the outcome is explained because it is logically derived from this law and the specification of some initial conditions. The empirical regularity is here in position of the *explanans*. However, Alice could keep on asking questions and say: *Why do stones tend to fall when they are dropped from a certain height?* In this case, the empirical regularity becomes what is to be explained, i.e. the *explanandum*, and calls for a more fundamental explanation. The latter is well given by a theoretical model of fall in the Newtonian framework, for instance. This explanation still rests on a law, yet here the law is not an empirical law but rather a fundamental principle.

We then distinguish two levels of scientific explanations: *practical* explanations and *fundamental* explanations. Practical explanations are those for which $EXPM = D$ and $EXPS = (\mathcal{E}, M_{emp})$. That is to say, the data are seen as explained by the empirical model. The second case, fundamental explanations, are those for which the empirical model is now what is to be explained, i.e. for which $EXPM = M_{emp}$ and $EXPS = (\mathcal{T}, M_{th})$. Both cases are summarized in figure 8.

3.2 Some preliminary clarifications: laws, explanations and predictions

From this perspective, we can already clarify some discussions and criticisms about the classical DN account of scientific explanations. We agree that explaining is deriving from a general (set of) law(s). However, the term "law" here should not be taken in the restrictive sense of physical laws, e.g. given under the form of an equation. A law can also consist in a general principle, like rationality principles in the social sciences. Also, as mentioned above, there appears to be two different kinds of laws: theoretical laws and empirical laws. Theoretical laws, in our framework, belong to the theoretical frame $\mathcal{T}$ and their explicative power is at work through some theoretical models. Empirical laws, on the contrary, are empirical regularities captured by empirical models. In the orignal DN model, a condition to be a good explanation is that the *explanans* be true. Obviously, this was originally thought in a logical-empiricist-like framework, within which the notion of truth has a precise meaning. Here, we would not demand a so restrictive constrain, but we can still discuss it in our framework. Indeed, for practical explanations, the demand that the explanans be "true" would correspond to the demand that the empirical model playing this role be well corroborated. In the case of fundamental explanations, this demand takes another form, as we know that fundamental laws and principles do not have exactly the same epistemological status as empirical laws. More precisely, from the Lakatos' research program viewpoint, fundamental principles, which play the role of explanans in fundamental explanations, are set as true or at least as defining the very framework in which we can express explanations. We leave here this discussion which actually goes beyond the scope of this article.

We do not completely agree, however, with the symmetry principle between explanation and prediction provided by the DN model.[7] In particular, in the case of a practical explanation (see figure 8) which takes a statistical form, as illustrated by the Coleman diagram, a genuine explanation rests on the claim of a causal relationship between variables at the macrolevel. However, even with a mere correlation, it is already possible to make some predictions. For instance, the chocolate consumption (in kg/year/capita) of a country is a quite good predictor of the rate of Nobel laureates in this country (in number per 10 million population) (Messerli 2012), even if there is probably nothing profound to be explained. Thus, these predictions are not explanations. In the same vein, from an empirical model it is possible to make predictions but if there is no data available at that time to compare with the predictions, then we cannot really talk about an explanation - because there is literally nothing to be explained yet.[8]

In the case of fundamental explanations, the same asymmetry seems to occur. Indeed, from a theoretical frame, we can build theoretical models from which we can make predictions, under the form of empirical models. Yet, if these empirical models are not supported by empirical data, i.e. if they are not corroborated (for instance, because there is no data available or because the data contradict the empirical model) then it is odd to talk about explanation, because there is, again, nothing to be explained.

[7]Let us precise that we take the term "prediction" in the wide sense which also includes "postdictions": it is a logical deduction from some hypothesis (which usually form a model) which can be ultimatelly compared to empirical data. Predictions usually correspond to the cases where data are not yet available at the time of the predictions (like the prediction of the light deviation by Einstein's theory of general relativity) while postdictions are the contrary (like the derivation of the Mercury perihelion advance (an empirical result known from the nineteenth century) from the same Einstein's theory in 1915).

[8]However, a practical explanation is a prediction, for it allows to derive some empirical results from a certain set of hypothesis. For the same reasons, a fundamental explanation is always a prediction.

3.3 Three kinds of inadequate ***explananda***

Given the general schema:

$$EXPS \xrightarrow{\text{Logically entails}} EXPM$$

criteria of a good (scientific) explanation can possibly bear on the explanans, on the explanandum or on the logical relationship between them. Our way of approaching the problem of scientific explanation, in this paper, is to focus on the explanandum, for it is not so common in the literature and still there are interesting directions to be explored. The question that we adress in this section is thus: what is a good scientific explanandum?

We base this work on the framework developed in the previous section and upon the different distinctions that have been made. We choose to tackle this question *negatively* by identifying (and exemplifying) three kinds of misleading or inadequate explananda.[9] What we mean here is identifying some typical situations for which the problem lies in the fact that this is not the right explanandum which is under study. The first category is when the basic variables in the empirical frame are actually not (operationally) defined independently. This leads to a solution of the well-known Empire State Building counter-example to the DN model. The second one contains empirical regularities which actually reduce to artefacts of the empirical frame, i.e. artefacts of the language at use to express them (and thus which are not genuine explananda for there is nothing deep to be explained). The third category is made of cases when an empirical model is defended whereas it is not supported by robust empirical evidence, and an alternative model exist and enjoys a perfectly acceptable explanation. We will take the examples of correlations without causality and the case of the empirical test of a drug. It will allow us to discuss in particular two counter-examples to the DN model: the barometer-and-storm story and the case of a uteroless person taking a pill which prevents pregnancy, and see how these apparent paradoxes vanish in our framework. Notice that the distinction made in the last section between practical and fundamental explanations is worth keeping in mind here: the first category apply to practical explanation, while the second and third one apply to fundamental ones.

3.3.1 Dependency relationships within the empirical frame

A well known counter-example to the DN model of explanation[10] is that of the Empire State Building's shadow. It is a case of practical explanation. During a sunny day, it is possible to derive the height of a building from the measure of its shadow. Indeed, there is a simple geometrical relationship between the length of the shadow L, the height of the building H and the angle α between the sun beams and the ground:

[9] Without claiming to be exhaustive.

[10] As already mentioned in the introduction, all typical counter-examples - this one as well as all the others cited in this paper - can be found summarized in (Salmon 1989) or Woodward & Ross (2021).

$$\tan(\alpha) = \frac{H}{L}. \tag{6}$$

Thus, we can measure the length of the shadow and the angle, and derive the height of the building. Or, conversely, if we know the height of the building, we can derive the length of the shadow from the same equation. Here, equation (6) plays the role of the empirical model M_{emp}, and the empirical frame is $\mathcal{E} = \{L, \alpha, H\}$. The problem is the following: deriving the length of the shadow L from the height of the building H can be identified with an explanatory move: the height of the building does explain, in some sense, the length of the shadow. However, the converse is not true, and it is puzzling. Indeed, it is quite counter-intuitive to say that the length of the shadow explains the height of the building, while this has exactly the same apparent logical form as the former. Thus, for some reasons that we have to enlighten, the length of the shadow can be an *explandum* in this model, but the height of the building cannot.

Let us present two similar cases. If an object is dropped from an initial height h_0, it takes t_f to reach the ground. One then gets the following relationship:

$$t_f^2 = \frac{2h_0}{g} \tag{7}$$

like in section 2.1.2. Thus, using equation (7), h_0 can be derived from the measure of t_f, and conversely t_f can be derived from the measure of h_0. However, only the second one can be identified with an explicative move. Indeed, it seems fair to claim that the value of the initial height explains in some sense that of the fall time, and it seems odd to hold the contrary, namely that the fall time could explain the initial height.

A last example is that of the Kepler's laws relating the mass M of a planet, the mass m of a satellite of it, the demi-axis a and the period T of the elliptic orbit of the satelit around the planet:

$$\frac{a^3}{T^2} = \frac{G(M+m)}{4\pi^2}. \tag{8}$$

This equation is sometimes used to derive the mass of some objects by measuring some features of objects gravitating around them. Similarly, we would not say that the mass M of the planet is explained by the period T of the trajectory, whereas the converse is true.

Consider now n moles of a gas at low pressure into an hermetic box of volume V. We can measure the pressure P and the temperature T. All variables are related by the ideal gas law:

$$PV = nRT. \tag{9}$$

Imagine that we change the temperature by ΔT and then we measure a change of pressure of ΔP, such that $\Delta P = nR\Delta T/V$. Or conversely, we change the pressure and observe a change of temperature which follows exactly the same law. In this case, however, it is quite natural to say that both moves are explanatory: the change of temperature does explain the change of pressure in the first configuration, and the change of pressure does explain the change of temperature in the second one, the explanation being made by the help of the equation (9).

All these examples are cases of practical explanations, i.e. explanations from an empirical model, without regard for the possible more fundamental explanation of these empirical laws. The puzzling question is the following: what is the difference between the three first cases, where there appears to be an asymmetry between logical derivation and explanation, and the fourth one, which is symmetric? Our answer rests on the analysis of the empirical frame $\mathcal{E}$ at work in each case. It turns out that in the three first cases, *the basic variables are not independent in their operationalized definition.* Indeed, operationalizing H, i.e. the height of the building, only needs to use the usual operationalized notion of distance. Yet, operationalizing L, i.e. the length of the shadow, needs the same notion of distance *and* the notion of shadow, which depends on the existence of the building, and in particular on the fact that it has a certain height. Thus, in the operationalized definition of L, the definition of H is implicitly included, whereas the contrary is not true. That is why the explanatory move is not symmetric.

The second case works the same: the initial height h_0 of the fall can be operationalized independently of the fall itself, whereas the fall time t_f, to be operationalized, needs the operationalization of duration *and* the fact that something falls from a certain height. Again, the definition of t_f depends implicitely on the definition of h_0 whereas the contrary is not true. The last example of satellites around a planet shows the same structure: masses m and M do not depend, in their respective definition, of whether one is gravitating around the other, while the variables a and T do depend on the existence of an elliptic orbit. Thus, a and T implicitely contain, in their respective operationalized definition, the existence of a gravitying system which features m and M, and thus only the latter can explain the former, not the contrary.

These are examples of a dependency relationship within the empirical frame. In the last example of the ideal gas, there is no such dependencies for T and P are operationalized as two *a priori* independent features (among others) of a gas. That is why in this case, it does not seem odd to say that both the variation of P or of T can explain the variation of the other (still *via* the ideal gas law).

3.3.2 Endogeneous regularities

A second kind of inadequate explanandum are endogeneous regularities, i.e. empirical regularities which are produced by the empirical framework and reduce to language artefacts. They can be analytical statements, i.e. statements which are true by virtue of the definitions of the terms they are made of, or artefacts due to the modelization hypothesis at work. These are somehow connected to the notion of unfalsifiable statements in the popperian sense (Popper 1959), i.e. statements which are empirically adequate but the adequacy of which do not teach us anything about the relevance of our theory. Let us give some concrete examples.

Simple analytical statement: poverty rate. In the social sciences as political science or economics, there are several different ways of defining poverty. Let us take one of them to illustrate our point on a first simple example: a person is qualified as poor if he/she earns less than 60% of the median income. Then, the poverty rate r_{poor} of a given country is the ratio between the number of people under the poverty threshold and the total population of reference. Then, we can state that in any country when this rate r_{poor} can be measured, we will always observe that $r_{poor} \leq 50\%$. Yet, there is nothing to be explained here, for this regularity is a mathematical theorem that we can derive from the mere definitions of the terms.

Cobb-Douglas production function. In neoclassical economic theories, an important concept is that of the aggregated production function, which relates the quantity of economic output (like the global production of a country) with a certain set of economic inputs (like labor, capital, ...) In particular, Paul Douglas and Charles Cobb proposed in 1928 such a function, relating the total production Q in a year with total amount of labor L and capital K measured in homogeneous units:

$$Q = AL^{\alpha}K^{\beta}, \tag{10}$$

where A is a parameter. They tested this function econometrically and found a very strong empirical adequation with data. In particular, they empirically found that the parameters of this model are related by $\alpha + \beta = 1$, which constitutes an argument supporting some basic hypothesis of neoclassical economics. However, some criticism have raised over time, and it has been recently claimed (Felipe & McCombie 2013) that actually, this relation reduces to an *accounting identity*. In other words, the relation (10) together with $\alpha + \beta = 1$ can actually be mathematically deduced from an identity which is true by definition. This is a quite clear example of a regularity which is endogeneous to the language at use. This observation entails that there is nothing deep to be explained here, despite the great number of theoretical discussions which has taken place about this result.

Berkson statistical "paradox". In 1938, Joseph Berkson published a paper (Berkson 1946) in which he warned about statistical studies in epidemiology conducted from hospital data. The point is the following. Let us assume that we want to investigate at which point two diseases A and B are related, in this purpose we can use hospital data: we take people who came to the hospital for disease A, measure at which degree of severity they had A, and then see whether they had disease B and at which degree (and *vice versa*). We could observe, for instance that A and B are positively correlated: the presence of A at a certain degree of severity then predicts the presence of B at a certain degree; or A and B could be uncorrelated, which would mean that the presence of one of both cannot be predicted from the presence of the other; and so on.

Berkson noticed that actually, the fact to take *hospital* data, and not data from a representative sample of the general population, can generate spurious negative correlations. Let us assume that A and B are *not* correlated at all in the general population.

In this case, we could still measure a correlation in hospital data. The reason is the following: people which are coming to hospital for A or B come with a quite severe degree of the disease. Thus, people who have both A and B at a low degree just do not come to the hospital, and thus are underrepresented in the study. Thus, if a person has A at a severe degree, this person is more likely to have B at a low one – see figure 9.

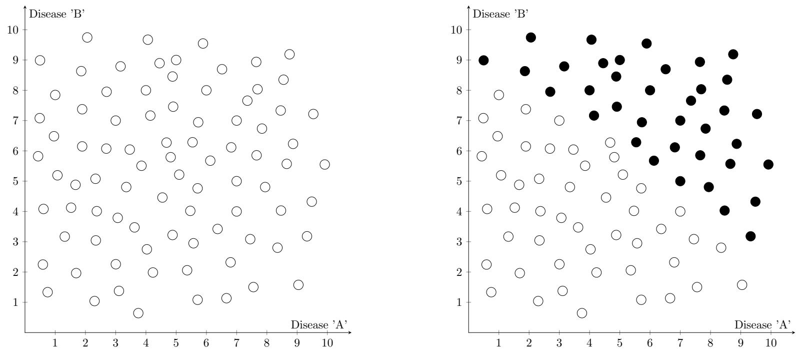

Figure 9: Berkson's statistical effect. Left: circles: general population distribution – no correlation. Right: filled circles: hospital data distribution – negative correlation.

Finally, the hospital data would make appear a negative correlation between A and B, but this correlation would be only due to the empirical frame at use. Taking this empirical correlation as something to be scientifically explained would then be a mistake. This biais is obviously not restricted to the domain of epidemiology but appears each time *the conditions to be part of a study are not independent from the variables we want to measure.* If the sample is not chosen with care, we can eventually end up trying to explain a negative correlation whereas there is actually nothing to explain.

Ptolemy's epicycloids. Another more subtle example is that of Ptolemy's geocentric planetary model, which was the predominant one over centuries until the late XVI's. In this model, the Earth is at the center of the universe and the other planets together with the sun are rotating around it. Some of these planets, however, are not following strict circles around the Earth, but more complicated trajectories called epicycloids. An epicycloid is generated by the rotation of a point along a circle the center of which is also rotating along another circle, and so on. It turns out that this model had very great empirical adequacy with observational data available at this time. Moreover, each time a discrepancy appeared, it was fixed by adding another epicycle, i.e. a new circle the center of which is rotating along the previous one. This model seems to be quite robust, since it can accomodate any anomaly without modifying its basic hypothesis.

However, the empirical adequacy of this model is actually a mathematical artefact. Any closed trajectory in the Earth's frame can actually be described by an epicycloid with a finite number of circles, given that the observational precision is itself finite - which is the case. This is because any such closed trajectory in two dimensions can be decomposed as the addition of two periodic functions, and that any periodic function can be approximated by its Fourier's decomposition,[11] i.e. a finite sum of cosine and sine functions. In two dimensions, this decomposition (which only rests on the mathematical modelization of trajectories as closed, continuous and derivable curves) exactly gives finite epicycloids.

It does not mean that this model cannot be used to make precise predictions, but just that the epicycloidal form is not something deep to be explained:[12] it reduces to a mathematical artefact due to the way these trajectories are modelized. On the contrary, Kepler's elliptical trajectories in the Sun's frame does say something deeper which appeals for a genuine explanation, for not any set of data can be described by such a model.

3.3.3 Spurious regularities

This third kind of inadequate *explanandum* has the following structure: an empirical model M_{emp} (which reproduces some data D) is assumed to call for an explanation, whereas there exists an alternative model M'_{emp} which:

- also reproduces the same data D,
- is well explained by a theoretical model within a well accepted theoretical frame.

Correlations without causation. A first example of this kind of inadequate *explanandum* is the case where D is the observation of a correlation between two variables A and B, and that:

- M_{emp} is the assumption of a causal relationship between A and B: $A \longrightarrow B$ which is taken as the *explanandum.*
- No explanatory mechanism (theoretical model) is known which could account for it.
- There exists a variable C such that both causal relationships $C \longrightarrow A$ and $C \longrightarrow B$ are well explained.

For instance, consider the positive correlation already mentioned between the Nobel laureates rate of a country (A) and the chocolate consumption (B) *per capita* (Messerli 2012). If a causal relationship (say $A \longrightarrow B$) is assumed, then one has to explain why consuming chocolate at a country scale increases the probability to earn Nobel prices, for instance by exhibiting a mechanism at the neuronal level describing how certain molecules chocolate contains act on cognitive abilities. However, another explanandum is available: this positive correlation is nothing but the reflect of the economical wealth (C) of the countries. The higher the wealth, the higher the chocolate consumption

[11] More precisely, for any closed, continuous and derivable function f, its Fourier series uniformally converges to f. That means that if a non null precision interval ϵ is given, the Fourier series of f will get at a distance less than ϵ from f with a finite number of terms.

[12] Exaclty like in the case of a non causal statistical correlation, as mentioned in section 3.2.

(for it is a luxury product) *and* the higher the Nobel prices laureates rates (for more ressources delegated to science in general). These two empirical models are then in competition while they do not call for the same kind of explanations: the former calls for a biological/neurological explanation, while the latter call for a sociological/economical one.

Another example which has the same structure is the other well-known counter-example of the classical DN model: the barometer and the storm. Consider a barometer: each time it shows a pressure drop, a storm occurs a bit of time later. It has the form of an empirical law, however it is odd to say that the pressure drop *explains* the apparition of the storm. From our viewpoint, it is just that it is not a genuine empirical model. Indeed, a correlation can make a prediction (but not interventionally) but is not a genuine basis for explanation: we need a causal relationship. Here, we have a temporal correlation between A (pressure drop) and B (storm), but assuming a causal relationship between them would be fallacious since another empirical model is available, explain the same data (the correlation) and rest on some more fundamental (mechanistic) explanations. Indeed, certain meteorological conditions C cause the pressure drop of the barometer ($C \longrightarrow A$) together with the apparition of the storm ($C \longrightarrow B$).

Specific effect of a medicine. Controlled randomized trials (CRT), e.g. to test the specific efficacy of a drug, are good examples where considerations coming from different disciplines have to be taken into account simultaneously. Indeed, testing the efficacy of a drug means comparing its effect to the effect of the administration of a placebo, i.e. a pill which looks like the medicine in all points except that it does not contain the substance the effect of which is being tested. It turns out that in general, the effect which is observed in the control group is higher that if nothing was given, even if this effect is much smaller than what has been believed for a long time (Hróbjartsson & Gøtzsche 2010). It is more and more established that the observed effect is actually the combination of several effects (called *contextual effects*): spontaneous recovery, submission to the doctor authority and conformity to his/her expectations, psychological conditioning and endogeneous release of painkillers (like endomorphins), and so on.

Facing some experimental data coming from a CRT, two possible empirical models are under test: the first one claims that there is no specific efficacy (i.e. that the entire effect can be reduced to contextual effects), while the second one claims that there is. However, in both cases there is something to be explained, but which does not call for the same theoretical frame.

A good example of an inadequate *explanandum* of this kind is the claim that homeopathic treatments have a specific efficacy. First, meta-analysis show that the more robust the methodology of the trial, the less the observed specific effect (Ernst 2002). Second, everyone, included homeopathy's defenders, agree that an homeopathic pill, due to the way it is produced, does not contain anything active - chimically, they are sugar pills. Thus, claiming that such pills could have a specific efficacy challenges our very knowledge in fundamental physics, chemistry and biology. However, the other model - that there is no specific effect - is perfectly explained by the contextual effects.

This example can be related to another well-known counter-example of the DN model: a uteroless person taking a pill which prevents pregnancy. Then, after taking this pill on a regular basis, this person indeed does not get pregnant. The general law: "a uteroless person taking this pill will not get pregnant" is perfectly true, however this does not constitute a genuine *explanandum.* This is exactly like the statement: "if you have a cold and take homeopathic pills, then your cold will disappear after few days". This is empirically confirmed aswell but misleading in its form for it implies that there is a causal relationship between the fact to take those pills and the fact to recover, while there is (probably) not. This example is then no more a counter-example in our framework, because from our viewpoint it merely does not constitute a genuine explanandum.

4 Conclusion

In this paper, we made a proposal for the general structure of scientific models:

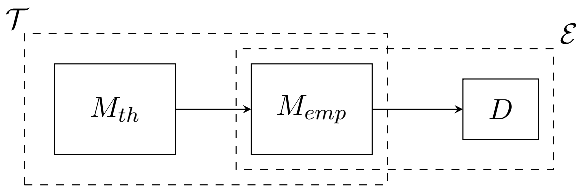

Figure 10: Meta-model of scientific models. $\mathcal{T}$ and $\mathcal{E}$ stand for the theoretical frame and the empirical frame, respectively. M_{th}, M_{emp} and D stand for the theoretical model, the empirical model and the data, respectively.

After having illustrated it on different examples from different scientific disciplines, we used it to clarify some discussions about scientific explanations. From this general structure, we identified two possible kinds of scientific explanations: practical explanations, for which $EXPS = (\mathcal{E}, M_{emp})$ and $EXPM = D$, and fundamental explanations, for which $EXPS = (\mathcal{T}, M_{th})$ and $EXPM = M_{emp}$. Their structure is thus basically that of the DN model: the explanation is made out of a (theoretical or empirical) *law.* There exist possibly several criteria for assessing the epistemological value of a scientific explanation, but we focused on *what is to be explained* i.e. on the explanandum. Our strategy was to enlighten three kinds of inadequate explananda, that is to say situations which may call for an explanation while there is actually nothing (or *another* thing) to be explained. In short, a good explanandum should respect at least these three aspects:[13]

- In the case of practical explanations:
  - The variation of a variable y may be explained by that of a variable x through a law relating them only if x does not depend, in its operationalized definition, from y. (Section 3.3.1.)
- In the case of fundamental explanations:

[13] Again, we do not claim to be exhaustive here.

- The empirical regularity must be a genuine regularity, and not a mere artefact of the empirical frame - that is to say, not an artefact of the very language in which this regularity is expressed. (Section 3.3.2.)
- For an empirically adequate empirical model to be a genuine explanandum, attention must be paid to other possible empirical models which would recover the same data *and* which would already benefit from a perfectly accepted explanation. (Section 3.3.3.)

The main result of this analysis, in addition to offering a general structure for scientific models, is that some common usual counter-examples to the DN model are neutralized within this unified framework. We nevertheless idenfity two main limits of this paper which has to be tackled in further investigations. The first one is the "proof" of the universality of such general structure for scientific models. Since the whole reasoning is based on this assumption, we need to support it as most as possible by exploring if it is indeed the case in a lot of different situations, in different scientific fields. The second direction in which further research has to be done is that of the assessment of scientific explanations not limited to explananda, including these reflections and using this general unified framework toward a more complete theory of scientific explanation.

References

Barberousse, A., Bonnay, D. & Cozic, M. (2011), *Précis de philosophie des sciences*, Vuibert.
URL: *https://halshs.archives-ouvertes.fr/halshs-00775660*

Berkson, J. (1946), 'Limitations of the application of fourfold table analysis to hospital data', *Biometrics* **2**(3), 47–53.
URL: *http://www.jstor.org/stable/3002000*

Bradford Hill, A. (1965), 'The environment and disease: Association or causation?', *Proceedings of the Royal Society of Medicine* **58**, 295–300.

Coleman, J. S. (1990), *Foundations of social theory*, Harvard University Press.

Ernst, E. (2002), 'A systematic review of systematic reviews of homeopathy', *British Journal of Clinical Pharmacology* **54**(6), 577–582.

Felipe, J. & McCombie, J. (2013), *The Aggregate Production Function and the Measurement of Technical Change: 'not Even Wrong'*, Edward Elgar.

Hedtröm, P. & Bearman, P. (2009), *The Oxford handbook of analytical sociology*, Oxford University Press.

Hempel, C. G. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: The Free Press.

Hempel, C. & Oppenheim, P. (1948), 'Studies in the logic of explanation', *Philosophy of Science* **15**, 135–175.

Howick, J., Glasziou, P. & Aronson, J. (2009), 'The evolution of evidence hierarchies: what can bradford hill's 'guidelines for causation' contribute?', *J R Soc Med* **102**, 186–194.

Hróbjartsson, A. & Gøtzsche, P. (2010), 'Placebo interventions for all clinical conditions', *Cochrane Database of Systematic Reviews* .

Manzo, G. (2014), *Analytical Sociology. Actions and Networks.*, Wiley series in computational and quantitative social science.

Manzo, G. (2021), *Research Handbook in Analytical Sociology*, Edward Edgar Publishing.

Messerli, F. H. (2012), 'Chocolate consumption, cognitive function, and nobel laureates', *New English Journal of Medicine* .

Offenbacher, M. (1901), *Konfession und soziale Schichtung, Eine Studie über die wirtschaftliche Lage der Katholiken und Protestanten in Baden*, Mohr.

Popper, K. (1959), *The Logic of Scientific Discovery*, Hutchinson and Co.

Quine, W. V. O. (1951), ‘Two dogmas of empiricism’, *Philosophical Review* **60**, 20–43.

Salmon, W. C. (1989), ‘4 decades of scientific explanation’, *Minnesota Studies in the Philosophy of Science* **13**, 3–219.

Weber, M. (1922), *Gesammelte Aufsätze zur Wissenschaftslehre*, J.C.B. Mohr.

Weber, M. (1965), *Essais sur la théorie de la science.*

Weber, M. (2001), *The protestant ethic and the spirit of capitalism*, Roxbury publishing company.

Woodward, J. & Ross, L. (2021), Scientific Explanation, *in* E. N. Zalta, ed., ‘The Stanford Encyclopedia of Philosophy’, Summer 2021 edn, Metaphysics Research Lab, Stanford University.

Ylikoski, P. (2021), *Understanding the Coleman boat*, Edward Elgar, pp. 49–63.